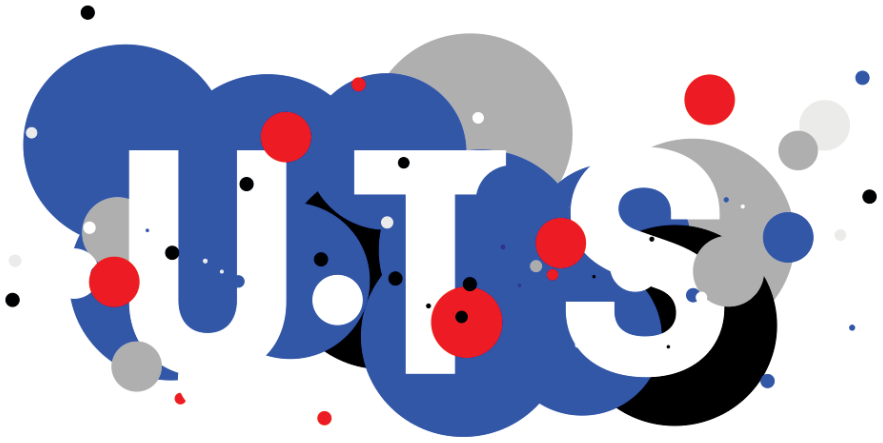

A Study of Multivariate Behavior and Anomaly Patterns:

Tensor Decomposition for Multiway Big Data

A Dissertation submitted for

The partial fulfillment of

Master of Engineering Research

By

Alina Rakhi Ajayan

University of Technology Sydney

New South Wales, Australia

2017

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as part of the collaborative doctoral degree and fully acknowledged within the text.

I also certify that I have written the thesis. Any help that I have received in my research work and the preparation of the dissertation itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

ACKNOWLEDGEMENT

My most profound gratitude is expressed to the University of Technology Sydney, and the School of Electrical and Data Engineering, FEIT, for this valuable opportunity to pursue impactful research. My attempts and motivation were primarily fuelled by my parents Mrs. Aleykutty Robert, Mr. G. R. Ajayan and my dear brother Akhil Minu Ajayan. I wish to express my gratitude to Dr. Zenon Chaczko, my advisor and supervisor through this two-year academic period.

The real drive and support came from my colleagues and friends at UTS – Manisha, Augustin, Hayat, Daniel, Firas, Ammar, Maria, Alaa, Mahmoud, Mohammad, Asma, Deepak, and Saha had been the constant voices in my head, egging me onward towards my goals and pursue my dreams. The entire Centre for Health Technologies Social group, starting with Dr. Adrian Bishop and Prof. Gyorgy Hutvagner had always been good mentors and advisors.

My candidature assessment panel members – the Hon'ble Chair and my Examiners, made my progress take a giant leap from the state of an in-depth Literature Review to a narrowed-down, streamlined research project. I hence extend my heartfelt gratitude to Dr. Daniel Franklin, Prof. Robin Braun, Dr. Negin Shariati Moghadam and Dr. Yi Cao, who brought me forward to the completion of this work.

Finally, but most importantly, the last few months of perseverance and dedication was hard to fight for and succeed. I could only push myself through the harsh write-up and implementation phase of research by the guidance of A/Prof. Raghav Menon and Michal Skowron.

Thanking you all from the bottom of my heart, for encouraging me to do my best, and to be at my best.

ABSTRACT

A vast majority of the today's information haul is through Cyber-Physical Systems (CPS). They represent the confluence of extensive data sets, tight time-constraints, latency issues and heterogeneous components. CPS architectures demand newer Big Data processing approaches. Typical applications span from the Internet-of-Things, across the World Wide Web to Smart Cities and Intelligent machines. A standard heterogeneous CPS installation, the Smart Energy Grid, is observed and the logistics are analyzed. The Smart Grid domain is weighed down by lack of unifying framework and systemic intelligence for autonomic management. Preliminary studies of the field under investigation shows how processing of Real-Time data, communication and control signaling is vital. Purely autonomic system governance is proven to be different from the contemporary definition. It takes the form of Interoperability (achieved through automation) instead of elemental Integration. That means autonomic (smart) management requires all elements to have fully controllable behavior.

This dissertation teste the hypothesis of applying Tensor decompositions and Factorizations - a momentum-gaining arithmetic tool - to this problem. The aim is to validate the prospects of higher order Anomaly Pattern Processing to capture intelligence along multiple modes of data flow. Tensorial Data representation captures information flows in Big Data, while Multivariate Anomaly Detection performs tracking of the time-series behavioral changes. Together, they implement Autonomic management in CPS super-systems. Uniqueness of this approach is highlighted by the novel multi-modal data flow imaging and models. Requirements of traditional anomalous event definition and cataloging in Data streams are removed. Tensor algebra is then studied for the scope of implementation concerning features, significance, and interpretation in terms of multi-modal data. Standard Decomposition rules and their derivatives, literature analysis on contemporary applications of Tensor algebra, and its scope on prominent real-world data processing problems are studied. Finally, the decomposition tool for Multi-way analysis is inferred, and proposed methodology is derived.

The Smart Grid Smart City Project commissioned by the Australian government is chosen as the data source investigated. The need for exhaustive examination of such repositories in the CPS Anomaly Detection context is also highlighted. Experimentation is done by applying Tensor Decomposition on the data set after normalization and pre-processing. Details of those phases, as well as the choice of coding platforms, the design of experimental frameworks, timelines estimated, and testing operations, are included in this work. The outcomes are the defined patterns extracted and their analysis-interpretation defended by proofs from actual events of the Project Trial phase.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES